International Journal of Public Health Excellence (IJPHE)
Vol. 5, Issue 2, January-May 2026, pp. 60~70
Journal Homepage: https://ejournal.ipinternasional.com/index.php/ijphe
ISSN: 2809-9826, DOI: 10.55299/ijphe.v5i2.1733

Classification of Lung Diseases Using The Naive Bayes Classifier Method at Madani Hospital Medan in 2025

Refi Sulistiasari[1], Handi Effendi[2], Agus Sumedi[3], Marlina Elfa Lubis[4], Zaim Anshari[5]
[1, 2, 3, 5] Lecturer of the Faculty of Medicine, Islamic University of North Sumatra, Jl. STM No. 77, Medan, Indonesia
[4] Lecturer at Mitra Sejati Health College, Jl. M. Basir No. 16 Pangkalan Mansyur Medan Johor North Sumatra

Article Info

Article history:

Received December 03, 2025
Revised January 06, 2026
Accepted February 05, 2026

Corresponding Author:

Refi Sulistiasari

Lecturer of the Faculty of Medicine, Islamic University of North Sumatra
Email:
refi.sulistiasari@fk.uisu.ac.id

ABSTRACT

The lungs are one of the human organs that are very important in the respiration process. There are several types of lung diseases, including Asthma, Bronchitis, Dyspnea Pleural Effusion, Empyema, Emphysema, Pulmonary Fibrosis, Lung Cancer, Interstitial Lung Disease, Pleurisy, Pneumonia, COPD, Tuberculosis. In this case there are difficulties in the process of classifying lung diseases, because the symptoms shown by sufferers of lung disease have similarities between one lung disease and another. The purpose of this study is to classify lung diseases using the Naive Bayes Classifier method. This method was chosen because it only requires a little training data to determine the estimated parameters needed in the classification process. This research was conducted at Madani General Hospital, Medan Area District, Medan City, North Sumatra, from February 1, 2025 to April 1, 2025. The data taken were medical records of lung disease patients from January 1, 2025 to February 1, 2025, totaling 134 patient data containing 19 disease symptoms and 6 disease diagnoses. From the test results using the Rapidminer application and data separation in the form of 34 test data and 100 training data with a data ratio of 7:3, an accuracy value of 97.06 was obtained.

Keywords: Lung Disease, Method, Classification

1. INTRODUCTION

The lungs are organs with respiratory functions. Respiration is the process of taking in and releasing energy stored in energy source substances through a chemical process that uses O2, the process of taking O2 to break down organic compounds into CO2, H2O and energy [1][2][3]. Lung disease is a disease that often occurs in humans and is usually caused by inhalation of polluted air, whether from dust, smoke, viruses, bacteria, fungi, or parasites that cause infections in the respiratory tract [4][5][6]. Infants, children, adolescents, adults, parents or the elderly can be infected or experience lung disease [7][8][9]. It is important to know the types of lung disease according to cases of lung disease in Indonesia, especially in the city of Medan, and the clinical symptoms experienced by patients with lung disease, one of the deadly lung diseases is tuberculosis [10][11]. According to the World Health Organization (WHO), tuberculosis is estimated to be the third leading cause of death in 2024, and the number of disabilities due to tuberculosis is projected to increase from 5th to 3rd in 2025 [12][13]. Data mining is the process of searching for rules or patterns from a large amount of data that haves a very high probability of finding new information. Data mining is often referred to as knowledge discovery database (KDD) because it is primarily used to find information from large-scale databases [14][15]. Classification is an important technique for data mining. Classification is the process of finding a pattern that describes or distinguishes each class of data with the aim of inferring the class of objects whose purpose is unknown [16][17]. People only know lung disease as asthma, but there are various types of lung diseases such as asthma, bronchitis,Dyspnea, pleural effusion, empyema, emphysema, pulmonary fibrosis, lung cancer, interstitial lung disease, pleurisy, pneumonia, pneumothorax, obstructive pulmonary disease, tuberculosis [18][19]. The lung diseases mentioned above certainly have different clinical symptoms and treatments depending on the type of lung disease experienced by the patient, therefore, lung disease sufferers must know the disease they are suffering from [20][21].

2. METHOD

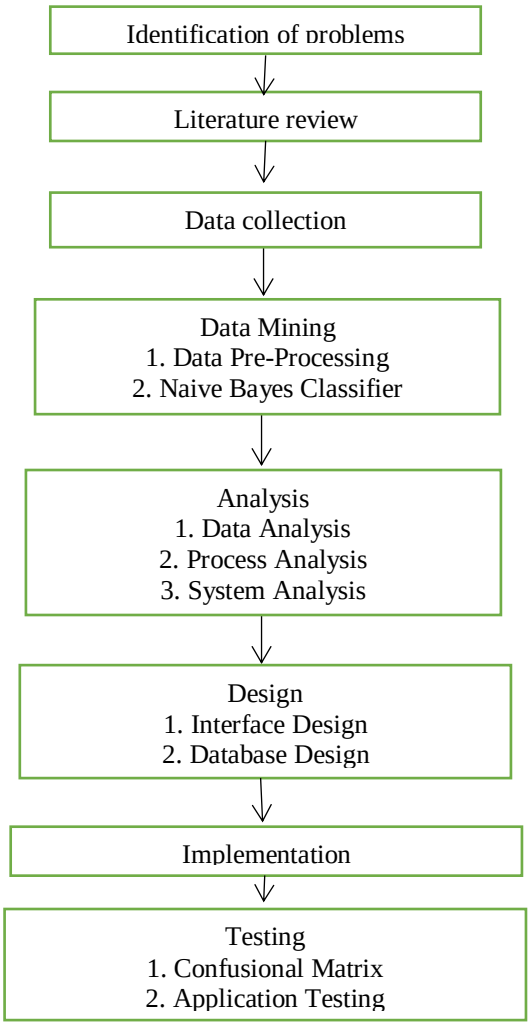
The Naïve Bayes Classifier method was used. The Naïve Bayes Classifier method is a simple probability classifier based on Bayes' Theorem [22][23]. The Naïve Bayes Classifier is a supervised learning algorithm, and this method assumes that the effect of certain features is independent of other features, which is why it is also called the naïve method [24][25].

The research steps are shown in Figure 1.

Figure 1 shows the stages of conducting research in sequence, namely,

1. Problem Identification: The purpose of this phase is to obtain an overview of existing problems.
2. At this stage, data and information related to the research are collected through various media, such as journals, books, the Internet, and related research.
3. Data Collection The data collection process was carried out during research at Madani Hospital, Medan Area District, Medan City, from February 1, 2025, to April 1, 2025. The data obtained were in the form of the medical records of patients with lung disease.
4. Data Mining: At stage, several pre-processing steps were carried out. Pre-processing is the process of converting or processing unorganized data into more organized data.
5. The purpose of pre-processing is to minimize failures, inconsistent data, or duplicate data.
6. Analysis: The purpose of the analysis was to align the data, display, and process of the research.
7. Design: This stage carries out the design of the interface and the database. Interface design is a form of description of the application being built, and database design is a form of data in the application being built.
8. Implementation To run the designed system, adequate equipment is required for running and coding the system. The purpose of coding is to change the design created so that it can be recognized by a computer.
9. Testing: To verify the accuracy of the system in performing the calculations, tests were conducted to measure its accuracy and performance. Rapidminer was used as a tool. Rapidminer is an open-core data processing tool typically used in machine learning, text mining, data mining, and predictive analytics.

Figure 1. Research Flow



RESULTS AND DISCUSSION

Data Preparation

The patient medical records used in this study were collected at Madani General Hospital, Medan Area District, Medan City, North Sumatra. The total data obtained from February 1, 2025, to April 1, 2025, amounted to 134 data points with 24 attributes and one class.

Data Pre-Processing

Before entering the data mining process, cleaning and transformation were performed on the data will be processed. The form of change from the Data Transformation is presented in table 1.

Table 1. Disease Data

Binominal	Attributes
Yes	1
No	0

The form of the disease symptom attributes afterundergoing the Data Transformation process in displayed in table 2.

Table 2. Disease Symptom Data

No	Symptom Code	Symptoms
1	G1	Shortness of breath
2	G2	Cough
3	G3	Coughing up blood
4	G4	Cough with phlegm
5	G5	Fever
6	G6	Weak
7	G7	Decreased Appetite
8	G8	Nausea
9	G9	Vomiting
10	G10	Chapter
11	G11	BAK
12	G12	History of Asthma
13	G13	History of Stroke
14	G14	History of TBC
15	G15	Headache
16	G16	Heartburn
17	G17	Chest Pain
18	G18	Weight Loss
19	G19	Night Sweats

The form of the disease name will be shown in table 3

Table 3. Disease Data

No	Symptom Code	Symptoms
1	D1	Asthma
2	D2	Bronchitis
3	D3	Dyspnea
4	D4	Pneumonia
5	D5	COPD
6	D6	TBC

The naive Bayes classifier method is a simple probability classifier based on Bayes' theorem. The Naïve Bayes Classifier is a supervised learning algorithm, and this method has a very strong assumption about the independence of each condition or event, which is why it is also called the Naïve Bayes Classifier method. The steps in the calculation process of the Naïve Bayes Classifier method are as follows:

a. Find the prior value for each class by calculating the class average.

$$P = \frac{X}{A}$$

b. Find the likelihood value for each class.

$$L = \frac{F}{B}$$

c. Find the posterior value of each class.

$P(c) \prod P(a|c)$

d. Find the confidence value of each class.

Confidence (c) P (c)

$\sum (c)$ x 100

Before carrying out the calculation process, the patient medical record data obtained will be divided into two datasets, namely, 100 training data and 34 testing data. The data that have gone through the pre-processing stage and are separated are displayed in table 4.

Table 4. Disease Data

Desc	Total Data	Training Data	Testing Data
Asthma	24	17	7
Bronchitis	11	8	3
Dyspnea	23	18	5
Pneumonia	26	19	7
COPD	31	23	8
TBC	19	15	4

Six classifications were formed in the output column: Asthma, Bronchitis, Dyspnea, Pneumonia, COPD, and Tuberculosis. The following explains the calculation process using the Naïve Bayes Classifier method with the following example data.

Table 5. Example of Test Data

No	Input	Symptoms
1	Symptoms of Shortness Breath	1
2	Cough Symptoms	1
3	Symptoms of Coughing Up Bloods	0
4	Phlegm Symptoms	1
5	Fever Symptoms	1
6	Symptoms of Weakness	0
7	Symptoms of Reduced Appetite	0
8	Nausea Symptoms	1
9	Symptoms of Vomiting	0
10	Symptoms of Defecate	0
11	Symptoms of Urination	0
12	Symptoms of Asthma	1
13	Symptoms of Stroke	0
14	Symptoms of TBC	0
15	Headache Symptoms	1
16	Symptoms of Heartburn	0
17	Chest Pain Symptoms	0
18	Symptoms of Weight Loss	0
19	Night Sweats Symptoms	0

Based on the test data presented in Table 5, the manual calculation process for one of the diagnoses using the Naïve Bayes Classifier algorithm was as follows:
Solution

Odds of Asthma (Asthma / Yes): 0.170
G1 Chance of Shortness of Breath (Yes / Shortness of breathBreath): 1
G2 Probability of Coughing (Yes / Coughing): 1
G3 Probability of coughing up blood (no / coughing up blood): 1
G4 Probability of Phlegm (Yes / Phlegm): 0.588
G5 Probability of Fever (Yes / Fever): 0.470
G6 Probability of Weakness (No|Weakness): 0.882
G7 Chances of Decreased Appetite (No|Decreased Appetite): 0.882
G8 Chance of Nausea (Yes|Nausea): 0.352
G9 Probability of Vomiting (No|Vomiting) :1
G10 Chance of defecation (No|defecation):
G11 Chance of BAK (No|BAK): 1
G12 Chances of a History of Asthma (Yes|History of Asthma): 0.529

G13 Chance of a History of Stroke (No|History of Stroke): 1
G14 Chance of History of Tuberculosis (No|History of Tuberculosis): 1
G15 Probability of Headache (Yes|Headache): 0.176
G16 Chance of Heartburn (No|Heartburn): 1
G17 Probability of Chest Pain (No|Chest Pain): 0.882
G18 Chance of Losing Weight (No|Weight Loss): 1
G19 Probability of Night Sweats (No|Night Sweats): 1
Join Probability Distribution (Yes|Asthma) Calculate Asthma(Asthma|newdata) = Asthma (Asthma|Yes) * Shortness of Breath (Yes|Shortness of Breath) * Cough (Yes|Cough) * Coughing Up Blood (Yes|Coughing Up Blood) * Phlegm (Yes|Phlegm)
* Fever (Yes|Fever) * Weakness (Yes|Weakness) * Decreased Appetite (Yes|Decreased Appetite) * Nausea (Yes|Nausea) * Vomiting (Yes|Vomiting) * Defecation (Yes|Defecation) * Urination (Yes|Urination) * History of Asthma
(Yes|History of Asthma) * History of Stroke (Yes|History of Stroke) * History of Tuberculosis (Yes|History of Tuberculosis) * Headache (Yes|Headache) * Heartburn (Yes|Heartburn) * Chest Pain (Yes|Chest Pain) * Weight Loss (Yes|Weight Loss) * Night Sweats (Yes|Night Sweats) = 0.170 * 1 * 1 * 1 * 0.588 * 0.470 * 0.882 * 0.882 * 0.352 * 1 * 1 * 1 * 0.529 * 1 * 1 * 0.176 * 1 * 0.882 * 1 * 1 = 0.00106595 or 0.11%.
From the calculation results, the same probability value was obtained for each disease diagnosis, as listed in table 6.

Table 6. Calculation Result

Asthma	Bronchitis	Dyspnea	Pneumonia	COPD	Tuberculosis
0.001	0.0001	0.00001	0.00006	0.00001	0.000001
6596	00000	025400	353	1296	00

The probability value obtained from each disease diagnosis was calculated to obtain the confidence value = 0.00106595 + 0 + 0.00002540 + 0.00006353 + 0.00001296 + 0 = 0.00116784.
After adding all the probability values, the confidence value can be calculated. The confidence value is calculated as follows:
Asthma Confidence Count = 0.00106595 = 0.00116784 = 0.9128
The same calculation was performed for each probability value of each disease diagnosis. From each diagnosis and the highest confidence value, possible classifications were determined using the Naïve Bayes Classifier algorithm. The confidence values are listed in Table 7.

Table 7. Example of Test Data

Asthma	Bronchitis	Dyspnea	Pneumonia	COPD	Tuberculosis
0.9128	0.0001	0.0217	0.0544	0.0111	0.0001

Classification Using Rapid Miner. The results of the Rapid Miner process are as follows. In Figure 2, several components are used, namely, data retrieval, naïve Bayes, application model, and performance. Retrieve functions as input data in the form of an excel, separated into two forms: Retrieve Data Training, which functions as training data, and Retrieve Data Testing, which functions as test data. Naïve Bayes functions as a calculation algorithm used in the mining process and is connected to the Apply Model, which connects retrieval data testing with naïve Bayes and retrieves data training. Performance functions to display the calculation results using the Naïve Bayes Classifier algorithm.

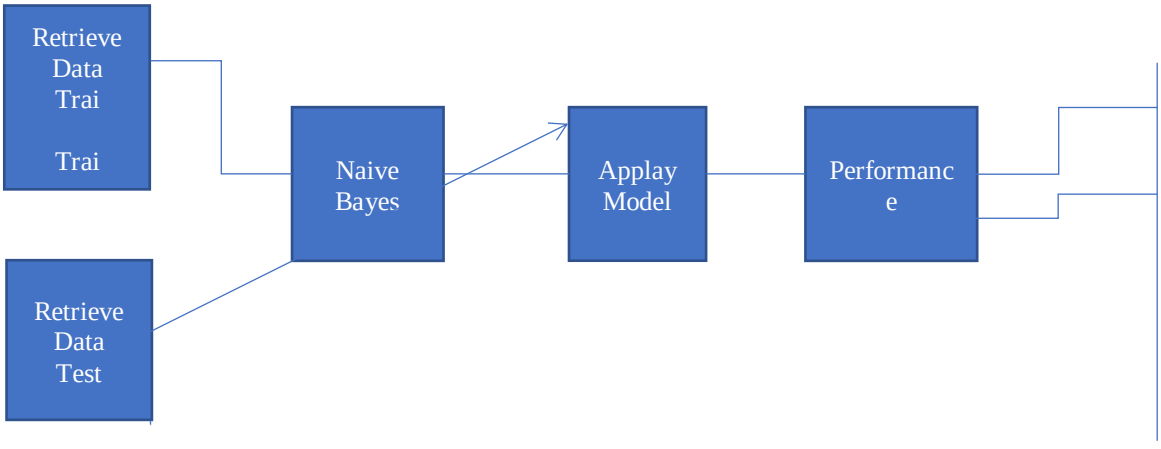


Figure 2. Naïve Bayes Classifier Calculation Series

For the data used, namely, 34 testing data and 100 training data, there are 19 attributes that are symptoms of a lung disease diagnosis, namely Shortness of Breath, Cough, Coughing Up Blood, Phlegm, Fever, Weakness, Decreased Appetite, Nausea, Vomiting, Defecation, Urination, History of Asthma, History of asthma.
Stroke, History of Tuberculosis, Headache, Heartburn, Chest Pain, Weight Loss, and Night Sweats.

Diagnosis	Out of breath	Cough	Coughing up blood	Phlegmy	Fever	Weak
Dyspnea	0	1	0	1	1	0
Tuberculosis	1	1	1	0	0	0
Asthma	1	1	0	0	1	0
Bronchitis	1	1	0	0	0	0
Tuberculosis	1	1	1	0	0	1
Tuberculosis	1	0	0	0	1	1
COPD	1	1	0	0	0	0
Asthma	1	1	0	1	0	0
Pneumonia	1	1	0	0	1	0
COPD	1	0	0	1	1	0

Figure 3. Data Attributes

System Design

A on the login page, a portal is displayed that directs users to log in when using the system, which can be seen in Figure 4.

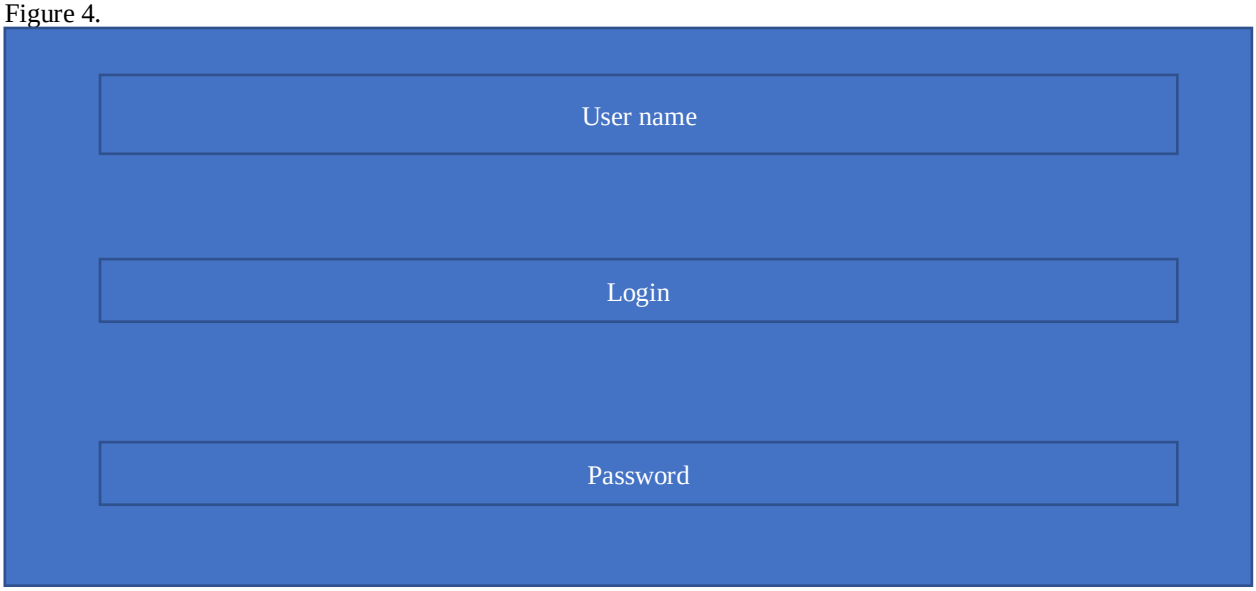


Figure 4. Login Page

b. Home Page

On the Main Page, the main part of the system is displayed, which contains several menu items such as the Main Page, Testing Data, Training Data, Naïve Bayes Process, Classification, and Logout, as shown in Figure 5.

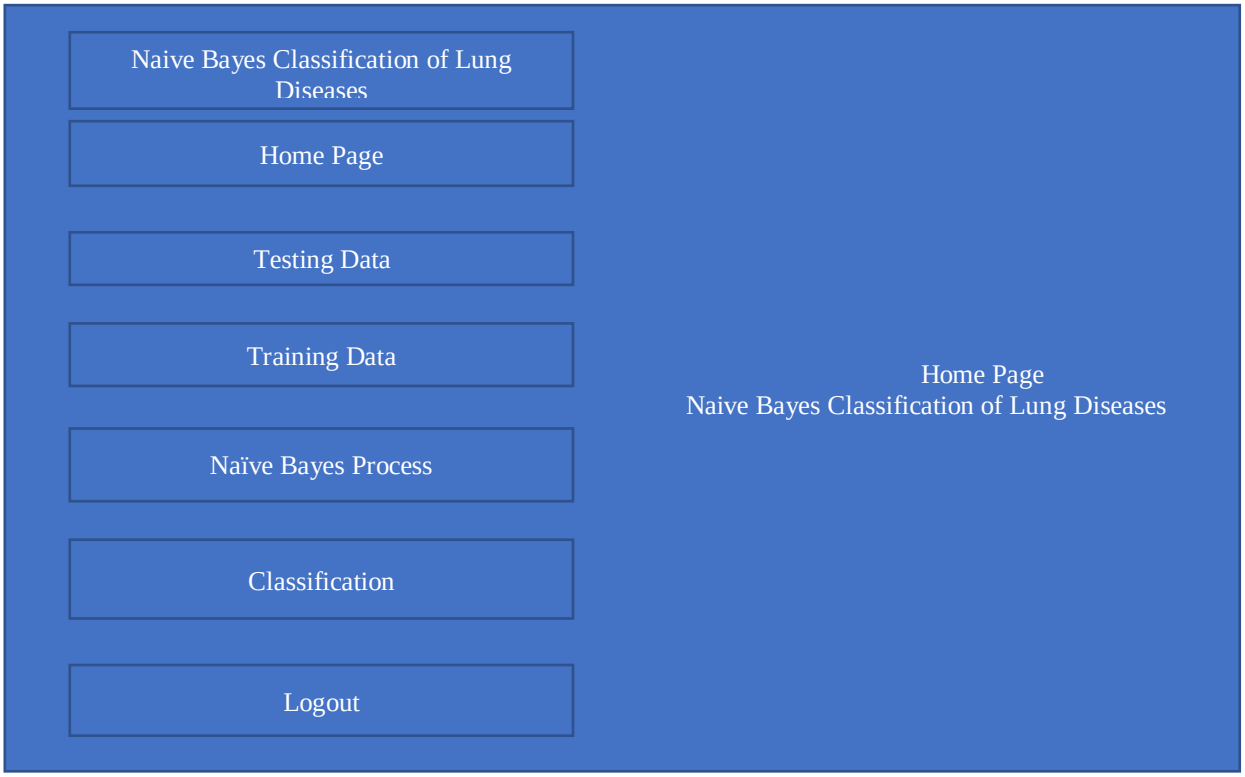


Figure 5. Main Page

c. Testing Data Page

On the Testing Data Page, a set of data is displayed that functions as the training data.
the calculation process can be seen in figure 6.

No	Name	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	Q 11	Q 12	Q 13	Q 14	Q 15	Q 16	Q 17	Q 18	Q 19	D P
1	Budi	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	TB
2	Amir	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Dy
3	Bayu	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	US
4	Umar	1	1	0	1	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	Br
5	Danu	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	Pk
6	John	1	1	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	Pn
7	Roni	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Dy
8	Gana	1	1	0	1	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	Br

Figure 6. Testing Data Page

d. Training Data Page

On the Training Data Page, a set of data that functions as the test data in the process is displayed.
calculations can be seen in Figure 7.

No	Name	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	Q 11	Q 12	Q 13	Q 14	Q 15	Q 16	Q 17	Q 18	Q 19	D P
1	Budi	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	TB
2	Amir	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	1	Dy
3	Bayu	1	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	US
4	Umar	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Br
5	Danu	1	1	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	1	Pk
6	John	1	0	0	0	1	1	1	0	0	1	0	0	0	1	0	0	0	0	0	Pn
7	Roni	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Dy
8	Gana	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	Br

Figure 7. Training Data Page

e. Naïve Bayes Process Page

On the Naïve Bayes Process Page, the entire calculation process is displayed, starting from finding the prior, likelihood, posterior, and posterior on the training data, as shown in Figure 8.

Naive Bayes

1. Find the Prior Value of Class / Diagnosis (Training Data)

No	Code	Diagnosis	Amount	Prior
1	D1	Asthma	17/100	0.17
2	D2	Bronchitis	8/100	0.8
3	D3	Dyspnea	18/100	0.18
4	D4	Pneumonia	19/100	0.19
5	D5	COPD	23/100	0.23
6	D6	Tuberculosis	15/100	0.15
Total				1

2. Likelihood of Each Class (Training Data)
Training Dataset

No	Name	G 1	G 2	G 3	G 4	G 5	G 6	G 7	G 8	G 9	G 10	G 11	G 12	G 13	G 14	G 15	G 16	G 17	G 18	G 19	Dx
1	So	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	Dy
2	Radi	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	TB

Figure 8. Naïve Bayes Process Page

f. New Data Classification Page
On the New Data Classification Page, a standalone form can be used to carry out the classification process on the input data, which can be seen in Figure 9.
Naive Bayes Classification

No	Name of Disease Symptoms Experienced	Information
1	Out of breath	
2	Cough	
3	Cough with phlegm	
4	Phlegmy	
5	Fever	
6	Weak	
7	Decreased Appetite	
8	Nauseous	
9	Vomit	
10	Dizziness	

Figure 9. New Data Classification Page

Testing
a. Using Testing Data
The first test process used 34 test data sets. The purpose of the calculations using the test data was to calculate the accuracy using the Naïve Bayes Classifier. The results of calculations using this system are shown in Fig.
Figure 10.

No	Name	D1	D2	D3	D4	D5	D6	Diagnosis (actual)	Diagnosis (Prediction)	Results
1	Abdul	0.000943	0.000007	0.003728	0.000099	0.536889	0.458335	COPD	COPD	Correct
2	Adrian	0.081211	0.235845	0.533113	0.000001	0.148075	0.001756	Dyspnea	Dyspnea	Correct
3	Ani	0.185217	0.119531	0.567404	0.000000	0.127581	0.000267	Dyspnea	Dyspnea	Correct
4	Ayu	0.008230	0.000072	0.009565	0.921690	0.060430	0.000013	Pneumonia	Pneumonia	Correct
5	Azizah	0.323215	0.219019	0.297047	0.000004	0.160715	0.000000	Dyspnea	Dyspnea	Correct
6	Dainar	0.000463	0.000002	0.000593	0.924378	0.074563	0.000001	Pneumonia	Pneumonia	Correct
7	Dila	0.516974	0.087579	0.158733	0.165309	0.071406	0.000000	Asthma	Asthma	Wrong
8	Dimas	0.008230	0.000072	0.009565	0.921690	0.060430	0.000013	Pneumonia	Pneumonia	Correct
9	Dinda	0.008230	0.000072	0.009565	0.921690	0.060430	0.000013	Pneumonia	Pneumonia	Correct
10	Donni	0.001371	0.444600	0.230234	0.227838	0.095955	0.000003	Bronchitis	Bronchitis	Correct

Figure 10. Testing Using the System

In addition to conducting trials using the program, calculations were performed using Rapid Miner. The Naïve Bayes Classifier algorithm calculations in Rapid Miner, using 34 test datasets and 100 training datasets with a 7:3 data ratio, yielded an accuracy of 97.06%. Of the 34 test datasets, one resulted in incorrect classification. The test results are shown in Figure 11.

Accuracy: 97.06%

No	Information	True COPD	True Dyspnea	True Pneumonia
1	Pred.COPD	8	0	0
2	Pred.Dyspnea	0	4	0
3	Pred.Pneumonia	0	0	7
4	Pred.Asthma	0	1	0

Figure 11. Confusion Table

b. Using Fresh (New) Data

The second calculation process uses fresh data. After achieving accuracy using the test data, the data can be classified on a new data classification page. After completing the form, the system displayed the symptoms experienced, as shown in Figure 12.

User Prediction Results

Symptoms Experienced

No	Symptom	Yes No
1	Out of breath	Yes
2	Cough	Yes
3	Cough with phlegm	No
4	Phlegmy	No
5	Fever	No
6	Weak	Yes
7	Decreased Appetite	No
8	Nauseous	No
9	Vomit	Yes
10	CHAPTER	No
11	BAK	No
12	History of Asthma	No
13	History of Stroke	No
14	History of Tuberculosis	No
15	Headache	No
16	Heartburn	No
17	Chest Pain	No
18	Weight Loss	No
19	Night Sweats	No

Figure 12. Patient Symptoms

After entering the symptoms experienced, the system calculates and displays the closest predicted lung disease percentage from the six types available. The entered symptoms yielded the highest confidence value, with a score of 0.984, predicting COPD. A summary of the program calculations is shown in Figure 13.

No	Types of Disease	Posterior	Confidence
1	Asthma	0.000000	0.000
2	Bronchitis	0.000000	0.000
3	Dyspnea	0.000000	0.001
4	Pneumonia	0.000000	0.000
5	PPOKO	0.000431	0.984
6	Tuberculosis	0.000007	0.015

Figure 13. Classification Output

Based on calculations using the naive Bayes model, the highest confidence value obtained was 0.984, and it was found that the patient had a 98.36% probability of having COPD compared to six other types of lung disease

3. CONCLUSION

Provide a statement that what is expected, as stated in the "INTRODUCTION" section can ultimately result in "RESULTS AND DISCUSSION" section, so there is compatibility. Moreover, it can also be added the prospect of the development of research results and application prospects of further studies into the next (based on result and discussion).

ACKNOWLEDGEMENTS

The authors thank all people and institutions, and in most cases, the sponsor and financial support acknowledgments.

REFERENCES

[1] S. I. Majid and M. Endang Dian Setioningsih, ST, MT, Sumber, "'Monitoring BPM, Suhu dan Respirasi Tampil PC via Bluetooth dan Pengiriman Data via SMS (Monitoring BPM dan Pengiriman Data via SMS),'

Semin.Tugas Akhir, 2018," 2018.
[2] A. S. Mulyani, "'Pemanasan Global, Penyebab, Dampak dan Antisipasinya,' Artik. Pengabdi. Masy., pp. 1–27, 2021," 2021.
[3] and D. G. J. Abidin, F. Artauli Hasibuan, K. Kunci, P. Udara, "'Pengaruh dampak pencemaran udara terhadap kesehatan untuk menambah pemahaman masyarakat awam tentang bahaya dari polusi udara,' Pros. SNFUR-4, no. September, pp. 1–7, 2019," 2019.
[4] and J. I. A. S. Pratama, S. Safrizal, "'Sistem Pakar Mendiagnosa Penyakit Gangguan Pernafasan Oleh Asap Rokok Menggunakan Metode Dempster Shafer,' It (Informatic Tech. J., vol. 9, no. 1, p. 79, 2021," 2021, doi: 10.22303/it.9.1.2021.79- 88.
[5] D. Kurniawan, F. Eka Febriansyah, and A. Amidya, "Implementasi Pengujian Theorema Bayes Pada Studi Kasus Pemilihan Smartphone Android," *J. Komputasi*, vol. 6, no. 2, pp. 11–14, Oct. 2018, doi: 10.23960/komputasi.v6i2.1744.
[6] M. F. Rifai, H. Jatnika, and B. Valentino, "Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)," *PETIR*, vol. 12, no. 2, pp. 131–144, Sep. 2019, doi: 10.33322/petir.v12i2.471.
[7] and U. C. R. Wana Dwi Putra, E. Prasetyo, "'Klasifikasi Penyakit Paru-Paru Menggunakan Metode Naive Bayes Studi Kasus Rs.Pku Muhammadiyah Ujung Pangkah Gresik,' no. 0561, p. 740186, 2020, [Online]. Available: http://eprints.umg.ac.id/3669/.," 2020.
[8] B. M. Sapata and D. Juniati, "'Klasifikasi Penyakit Paru Berdasarkan Citra X-Ray Thorax Menggunakan Metode Fraktal Box Counting,' MATHunesa, vol. 7, no. 3, pp. 228–235, 2019," 2019.
[9] S. Suprianto, "Perbandingan Metode Naïve Bayes Classifier Dan Holistic Lexicon Based Dalam Analisis Sentimen Angket Mahasiswa," *JSI J. Sist. Inf.*, vol. 11, no. 2, Oct. 2019, doi: 10.36706/jsi.v11i2.9140.
[10] and W. A. A. S. J. S. Tyas, M. Febianah, F. Solikhah, A. L. Kamil, "'Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan,' J. Teknol. Inf. Dan Komun., vol. 8, no. 1, pp. 86–99, 2021," 2021.
[11] and Y. S. A. C. Adi, D. P. Lestari, F. S. S. Elsa, "'Online School Sentiment Analysis in Indonesia on Twitter Using The Naïve Bayes Classifier and Rapid Miner Tools,' Int. J. Innov. Sci. Res. Technol., vol. 71, no. 1, pp. 785–788, 2022," 2022.
[12] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 219, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.
[13] D. Novianti, "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 49–54, Mar. 2019, doi: 10.31294/p.v21i1.4979.
[14] Dimsyiar M Al Hafiz, Khoirul Amaly, Javen Jonathan, M Teranggono Rachmatullah, and Rosidi, "Sistem Prediksi Penyakit Jantung Menggunakan Metode Naive Bayes," *J. Rekayasa Elektro Sriwij.*, vol. 2, no. 2, pp. 151–157, May 2021, doi: 10.36706/jres.v2i2.29.
[15] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, Dec. 2018, doi: 10.26418/jp.v4i2.27526.
[16] R. Nofitri and J. Eska, "'Implementasi Data Mining Klasifikasi C4.5 Dalam Menentukan Kelayakan Pengambilan Kredit,' Semin. Nas. R., vol. 9986, no. September, pp. 307–310, 2018," 2018.
[17] D. T. Satya and N. Hidayat, "'Sistem Pakar Diagnosis Penyakit Sapi Ternak Potong Menggunakan Metode Naïve Bayes - Certainty Factor,' J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 2, no. 10, pp. 3406–3410, 2018," 2018.
[18] M. Iqbal et al., "'Implementasi Metode Certainty Factor dalam Sistem Pakar Diagnosis Penyakit Paru-Paru Berbasis Android,' Coding J. Komput. dan Apl., vol. 07, no. 03, pp. 155–164, 2019," 2019.
[19] H. H. Rahman, D. Niemann, and S. H. Munson-McGee, "Association between asthma, chronic bronchitis, emphysema, chronic obstructive pulmonary disease, and lung cancer in the US population," *Environ. Sci. Pollut. Res.*, vol. 30, no. 8, pp. 20147–20158, Oct. 2022, doi: 10.1007/s11356-022-23631-3.
[20] F. N. Azizah and D. Juniati, "ANALISIS JENIS PENYAKIT PARU-PARU BERDASARKAN CHEST X-RAY MENGGUNAKAN METODE FUZZY C-MEANS," *MATHunesa J. Ilm. Mat.*, vol. 9, no. 2, pp. 322–331, Aug. 2021, doi: 10.26740/mathunesa.v9n2.p322-331.
[21] Purnamasari, "Analisa Klasifikasi Loyalitas Siswa Lembaga Pendidikan Tari dengan Metode Naïve Bayes," *J. Ilm. Komputasi*, vol. 19, no. 1, Mar. 2020, doi: 10.32409/jikstik.19.1.157.
[22] Ajeng Mega Pratiwi and A. K. Mutaqin, "Penerapan Algoritma Naïve Bayes Classifier dalam Memprediksi Status Keberlanjutan Polis Nasabah Asuransi PT.X," *J. Ris. Stat.*, vol. 1, no. 2, pp. 117–126, Dec. 2021, doi: 10.29313/jrs.v1i2.435.
[23] Ariski Muhammad Nazmi, Prajoko, and A. Pambudi, "Classification Analysis Of The Eligibility Of Recipients Of Non-Cash Food Assistance And Family Hope Programs In The City Of Sukabumi Using The Naïve Bayes

Classifier Algorithm," *J. Artif. Intell. Eng. Appl.*, vol. 3, no. 2, pp. 546–550, Feb. 2024, doi: 10.59934/jaiea.v3i2.420.
[24] Q. Stier, J. Hoffmann, and M. C. Thrun, "Classifying with the Fine Structure of Distributions: Leveraging Distributional Information for Robust and Plausible Naïve Bayes," *Mach. Learn. Knowl. Extr.*, vol. 8, no. 1, p. 13, Jan. 2026, doi: 10.3390/make8010013.
[25] R. D. L. N. Karisma, "Classification Cyber Harassment on Twitter using Multinomial Naïve Bayes," *MATICS J. Ilmu Komput. dan Teknol. Inf. (Journal Comput. Sci. Inf. Technol.*, vol. 17, no. 2, pp. 52–62, Sep. 2025, doi: 10.18860/mat.v17i2.29866.